Capstone Project Proposal: Analyzing Hand Color for Health Status Prediction

Adli Karadsheh

July 21, 2025

Capstone 3

Introduction and Problem Statement

In healthcare, early detection of circulatory or oxygenation issues can significantly improve patient outcomes. One subtle indicator on physical exam of such problems is the color of a person's hands, where bluish hues may signal poor blood flow or low oxygen levels (hypoxia), while pink tones typically indicate healthy circulation. This project addresses the real-world problem of non-invasive health monitoring for at-risk individuals, such as those with cardiovascular conditions or respiratory disorders, by developing a machine learning model to classify hand images as "healthy" (pink tones) or "unhealthy" (bluish hues) based on color analysis.

The problem solution will utilize data science and machine learning approaches, leveraging image processing, feature extraction, and binary classification techniques. By automating this analysis, the model could be integrated into a mobile app for remote health screening, empowering users to seek timely medical advice.

Dataset and Data Sources

The dataset consists of hand images divided into two categories: "Healthy" and "Unhealthy." The "Healthy" folder will contain 50-100 images sourced from free stock photo websites (e.g., Pexels, Unsplash, Pixabay), or paid (Adobe Stock), featuring hands with normal pink skin tones. The "Unhealthy" folder includes approximately 50 images obtained in an outpatient clinic , of hands of persons with chronic conditions with bluish hues indicative of potential health issues, supplemented by additional images if needed.

To support the problem statement, images will be processed using MediaPipe Hands for landmark detection and RGB color extraction at 20 key points (excluding the wrist). This

would result in 60 features per image (20 points × 3 RGB channels). The dataset will be balanced using data augmentation techniques (e.g., rotation, brightness adjustments) via TensorFlow's ImageDataGenerator to mitigate imbalance and ensure robust training. All images are labeled binary: 1 for healthy, 0 for unhealthy.

Approach to Solving the Problem

The proposed solution follows a data science pipeline to classify hand health status:

- **Data Preprocessing**: Use OpenCV for image loading and MediaPipe Hands to detect landmarks and extract RGB values.

- **Feature Engineering**: Flatten the 60 RGB features for input into models.

- **Modeling**: Train a neural network with Keras (Sequential model with dense layers and sigmoid output) for binary classification. Additionally, use Random Forest for feature importance to identify key landmarks (e.g., fingertips) most indicative of health status.

- **Evaluation**: Assess using accuracy, loss (binary cross-entropy), ROC-AUC, and per-class loss to handle potential imbalance.

- **Deployment**: Save the model as HDF5 for potential app integration.

Problem Identification And the Data Science Method (DSM)

Following the DSM's Problem Identification step, the substeps are addressed as follows:

1.**Problem Definition**: The challenge is detecting early signs of health issues via hand color, which could reduce healthcare burdens by enabling self-monitoring.

2. **Stakeholders**: Primary stakeholders include patients (end-users), healthcare providers (for validation), and developers (for app integration).

3. **Goals and Success Criteria**: The goal is a model with >80% accuracy and AUC >0.85 on test data. Success is measured by reliable classification and identification of important features (e.g., via permutation importance).

4. **Risks and Constraints**: Risks include dataset imbalance (mitigated by augmentation) and variability in lighting/skin tones (addressed by diverse sourcing). Constraints are limited healthy images (supplemented from public sources) and computational.

5. **Scope and Feasibility**: The project is scoped to classification only, excluding real-time deployment. It is feasible within 1 week, using open-source tools and a modest dataset.

Conclusion

This project proposes a practical ML solution to a healthcare problem, supported by a curated dataset and a structured DSM approach. Future extensions could include multi-class health indicators or integration with wearable devices.

References: MediaPipe Documentation, TensorFlow Tutorials, Public Image Sources (Pexels, Unsplash).